



1600

## RAW SEQUENCE LISTING

DATE: 06/12/2003

PATENT APPLICATION: US/09/915,873A

TIME: 14:30:08

Input Set : A:\Seqlist.txt

Output Set: N:\CRF4\06122003\I915873A.raw

4 <110> APPLICANT: Stomp, Anne-Marie  
 5 Dickey, Lynn  
 6 Gasdaska, John  
 9 <120> TITLE OF INVENTION: Expression of Biologically Active  
 10 Polypeptides in Duckweed  
 12 <130> FILE REFERENCE: 40989/237225  
 14 <140> CURRENT APPLICATION NUMBER: 09/915,873A  
 15 <141> CURRENT FILING DATE: 2001-07-26  
 17 <150> PRIOR APPLICATION NUMBER: US 60/293,330  
 18 <151> PRIOR FILING DATE: 2001-05-23  
 20 <150> PRIOR APPLICATION NUMBER: US 60/221,705  
 21 <151> PRIOR FILING DATE: 2000-07-31  
 23 <160> NUMBER OF SEQ ID NOS: 12  
 25 <170> SOFTWARE: FastSEQ for Windows Version 4.0  
 27 <210> SEQ ID NO: 1  
 28 <211> LENGTH: 554  
 29 <212> TYPE: DNA  
 30 <213> ORGANISM: Zea mays  
 32 <400> SEQUENCE: 1  
 33 gatcaagtgc aaaggtccgc cttgtttctc ctctgtctct tgatctgact aatcttggtt 60  
 34 tatgattcgt tgagtaattt tggggaagc ttcgtccaca gttttttttt cgatgaacag 120  
 35 tgccgcagtg gcgctgatct tgtatgctat cctgcaatcg tggatgaact atgtctttta 180  
 36 tacccttcac taccatgaaa agactagtaa tctttctcga tgtaacatcg tccagcactg 240  
 37 ctattaccgt gtggtccatc cgacagtcgt gctgaacaca tcatacgata ttgagcaaag 300  
 38 atctatcttc cctgttcttt aatgaaagac gtcattttca tcagtatgat ctaagaatgt 360  
 39 tgcaacttgc aaggaggcgt ttctttcttt gaatttaact aactcgttga gtggccctgt 420  
 40 ttctcggacg taaggccttt gctgctccac acatgtccat tcgaatttta ccgtgttttag 480  
 41 caagggcgaa aagtttgcac cttgatgatt tagcttgact atgcgattgc tttcctggac 540  
 42 ccgtgcagct gcgg 554  
 44 <210> SEQ ID NO: 2  
 45 <211> LENGTH: 498  
 46 <212> TYPE: DNA  
 47 <213> ORGANISM: Artificial Sequence  
 49 <220> FEATURE:  
 50 <223> OTHER INFORMATION: Duckweed codon optimized nucleotide sequence  
 51 encoding human alpha-2B interferon  
 W--> 53 <221> NAME/KEY: CDS  
 54 <222> LOCATION: (1)...(498)  
 W--> 56 <400> 2  
 57 tgc gac ctc ccc cag acc cac agc ctc ggg tcc cgc cgc acc ctc atg 48  
 58 Cys Asp Leu Pro Gln Thr His Ser Leu Gly Ser Arg Arg Thr Leu Met  
 59 1 5 10 15  
 61 ctg ctg gcg cag atg cgc cgc atc tcg ctc ttc agc tgc ctg aag gac 96

ENTERED

## RAW SEQUENCE LISTING

DATE: 06/12/2003

PATENT APPLICATION: US/09/915,873A

TIME: 14:30:08

Input Set : A:\Seqlist.txt

Output Set: N:\CRF4\06122003\I915873A.raw

```

62 Leu Leu Ala Gln Met Arg Arg Ile Ser Leu Phe Ser Cys Leu Lys Asp
63          20          25          30
65 cgc cac gac ttc ggc ttc ccg cag gag gag ttc ggc aac cag ttc cag 144
66 Arg His Asp Phe Gly Phe Pro Gln Glu Glu Phe Gly Asn Gln Phe Gln
67          35          40          45
69 aag gcc gag acg atc ccc gtg ctc cac gag atg atc cag cag atc ttc 192
70 Lys Ala Glu Thr Ile Pro Val Leu His Glu Met Ile Gln Gln Ile Phe
71          50          55          60
73 aac ctg ttc agc acc aag gac agc tcg gcc gcc tgg gac gag acc ctg 240
74 Asn Leu Phe Ser Thr Lys Asp Ser Ser Ala Ala Trp Asp Glu Thr Leu
75 65          70          75          80
77 ctc gac aag ttc tac acc gag ctg tac cag cag ctc aac gac ctg gag 288
78 Leu Asp Lys Phe Tyr Thr Glu Leu Tyr Gln Gln Leu Asn Asp Leu Glu
79          85          90          95
81 gcg tgc gtg atc cag ggg gtt ggg gtt acg gag acg ccg ctg atg aag 336
82 Ala Cys Val Ile Gln Gly Val Gly Val Thr Glu Thr Pro Leu Met Lys
83          100          105          110
85 gag gac agc atc ctc gcc gtg cgc aag tac ttc cag cgc atc acg ctc 384
86 Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr Leu
87          115          120          125
89 tac ctc aag gag aag aag tac agc ccg tgc gcc tgg gag gtc gtt cgc 432
90 Tyr Leu Lys Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val Arg
91          130          135          140
93 gcc gag atc atg cgc tcc ttc agc ctg agc acc aac ctc cag gag agc 480
94 Ala Glu Ile Met Arg Ser Phe Ser Leu Ser Thr Asn Leu Gln Glu Ser
95 145          150          155          160
97 ctc cgc tcc aag gag taa 498
98 Leu Arg Ser Lys Glu *
99          165
102 <210> SEQ ID NO: 3
103 <211> LENGTH: 96
104 <212> TYPE: DNA
105 <213> ORGANISM: Oryza sativa
107 <400> SEQUENCE: 3
108 accatgcagg tcctgaacac gatgggtcaac aagcaacttcc tctccctgtc cgctctcatc 60
109 gtcctcctcg ggctgagcag caacctcacc gccggc 96
111 <210> SEQ ID NO: 4
112 <211> LENGTH: 188
113 <212> TYPE: PRT
114 <213> ORGANISM: Homo sapiens
116 <400> SEQUENCE: 4
117 Met Ala Leu Thr Phe Ala Leu Leu Val Ala Leu Leu Val Leu Ser Cys
118 1          5          10          15
119 Lys Ser Ser Cys Ser Val Gly Cys Asp Leu Pro Gln Thr His Ser Leu
120          20          25          30
121 Gly Ser Arg Arg Thr Leu Met Leu Leu Ala Gln Met Arg Arg Ile Ser
122          35          40          45
123 Leu Phe Ser Cys Leu Lys Asp Arg His Asp Phe Gly Phe Pro Gln Glu
124          50          55          60

```

## RAW SEQUENCE LISTING

DATE: 06/12/2003

PATENT APPLICATION: US/09/915,873A

TIME: 14:30:08

Input Set : A:\Seqlist.txt

Output Set: N:\CRF4\06122003\I915873A.raw

```

125 Glu Phe Gly Asn Gln Phe Gln Lys Ala Glu Thr Ile Pro Val Leu His
126 65 70 75 80
127 Glu Met Ile Gln Gln Ile Phe Asn Leu Phe Ser Thr Lys Asp Ser Ser
128 85 90 95
129 Ala Ala Trp Asp Glu Thr Leu Leu Asp Lys Phe Tyr Thr Glu Leu Tyr
130 100 105 110
131 Gln Gln Leu Asn Asp Leu Glu Ala Cys Val Ile Gln Gly Val Gly Val
132 115 120 125
133 Thr Glu Thr Pro Leu Met Lys Glu Asp Ser Ile Leu Ala Val Arg Lys
134 130 135 140
135 Tyr Phe Gln Arg Ile Thr Leu Tyr Leu Lys Glu Lys Lys Tyr Ser Pro
136 145 150 155 160
137 Cys Ala Trp Glu Val Val Arg Ala Glu Ile Met Arg Ser Phe Ser Leu
138 165 170 175
139 Ser Thr Asn Leu Gln Glu Ser Leu Arg Ser Lys Glu
140 180 185
143 <210> SEQ ID NO: 5
144 <211> LENGTH: 165
145 <212> TYPE: PRT
146 <213> ORGANISM: Homo sapiens
148 <400> SEQUENCE: 5
149 Cys Asp Leu Pro Gln Thr His Ser Leu Gly Ser Arg Arg Thr Leu Met
150 1 5 10 15
151 Leu Leu Ala Gln Met Arg Arg Ile Ser Leu Phe Ser Cys Leu Lys Asp
152 20 25 30
153 Arg His Asp Phe Gly Phe Pro Gln Glu Glu Phe Gly Asn Gln Phe Gln
154 35 40 45
155 Lys Ala Glu Thr Ile Pro Val Leu His Glu Met Ile Gln Gln Ile Phe
156 50 55 60
157 Asn Leu Phe Ser Thr Lys Asp Ser Ser Ala Ala Trp Asp Glu Thr Leu
158 65 70 75 80
159 Leu Asp Lys Phe Tyr Thr Glu Leu Tyr Gln Gln Leu Asn Asp Leu Glu
160 85 90 95
161 Ala Cys Val Ile Gln Gly Val Gly Val Thr Glu Thr Pro Leu Met Lys
162 100 105 110
163 Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr Leu
164 115 120 125
165 Tyr Leu Lys Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val Arg
166 130 135 140
167 Ala Glu Ile Met Arg Ser Phe Ser Leu Ser Thr Asn Leu Gln Glu Ser
168 145 150 155 160
169 Leu Arg Ser Lys Glu
170 165
173 <210> SEQ ID NO: 6
174 <211> LENGTH: 31
175 <212> TYPE: PRT
176 <213> ORGANISM: Oryza sativa
178 <400> SEQUENCE: 6
179 Met Gln Val Leu Asn Thr Met Val Asn Lys His Phe Leu Ser Leu Ser

```

## RAW SEQUENCE LISTING

DATE: 06/12/2003

PATENT APPLICATION: US/09/915,873A

TIME: 14:30:08

Input Set : A:\Seqlist.txt

Output Set: N:\CRF4\06122003\I915873A.raw

```

180 1          5          10          15
181 Val Leu Ile Val Leu Leu Gly Leu Ser Ser Asn Leu Thr Ala Gly
182          20          25          30
185 <210> SEQ ID NO: 7
186 <211> LENGTH: 31
187 <212> TYPE: PRT
188 <213> ORGANISM: Artificial Sequence
190 <220> FEATURE:
191 <223> OTHER INFORMATION: Modified rice alpha-amylase signal peptide
193 <400> SEQUENCE: 7
194 Met Gln Val Leu Asn Thr Met Val Asn Lys His Phe Leu Ser Leu Ser
195 1          5          10          15
196 Val Leu Ile Val Leu Thr Val Leu Ser Ser Asn Leu Thr Ala Gly
197          20          25          30
200 <210> SEQ ID NO: 8
201 <211> LENGTH: 21
202 <212> TYPE: PRT
203 <213> ORGANISM: Arabidopsis thaliana
205 <400> SEQUENCE: 8
206 Met Lys Thr Asn Leu Phe Leu Phe Leu Ile Phe Ser Leu Leu Leu Ser
207 1          5          10          15
208 Leu Ser Ser Ala Glu
209          20
212 <210> SEQ ID NO: 9
213 <211> LENGTH: 554
214 <212> TYPE: DNA
215 <213> ORGANISM: Zea mays
217 <400> SEQUENCE: 9
218 gatcaagtgc aaaggtccgc cttgtttctc ctctgtctct tgatctgact aatcttggtt 60
219 tatgattcgt tgagtaattt tggggaaagc ttcgtccaca gttttttttt cgatgaacag 120
220 tgccgcagtg gcgctgatct tgtatgctat cctgcaatcg tggatgaactt atgtctttta 180
221 tacccttcac taccatgaaa agactagtaa tctttctcga tgtaacatcg tccagcactg 240
222 ctattaccgt gtggtccatc cgacagtctg gctgaacaca tcatacgata ttgagcaaag 300
223 atctatcttc cctgtttctt aatgaaagac gtcattttca tcagtatgat ctaagaatgt 360
224 tgcaacttgc aaggaggcgt ttctttcttt gaatttaact aactcgttga gtggccctgt 420
225 ttctcggacg taaggccttt gctgctccac acatgtccat tcgaatttta ccgtgttttag 480
226 caagggcgaa aagtttgcac cttgatgatt tagcttgact atgcgattgc tttcctggac 540
227 ccgtgcagct gcgg                                     554
229 <210> SEQ ID NO: 10
230 <211> LENGTH: 498
231 <212> TYPE: DNA
232 <213> ORGANISM: Homo sapiens
234 <400> SEQUENCE: 10
235 tgtgatctgc ctcaaaccac cagcctgggt agcaggagga ccttgatgct cctggcacag 60
236 atgaggagaa tctctctttt ctctgcttg aaggacagac atgacttttg atttccccag 120
237 gaggagtttg gcaaccagtt ccaaaggct gaaaccatcc ctgtcctcca tgagatgatc 180
238 cagcagatct tcaatctctt cagcacaag gactcatctg ctgcttgga tgagaccctc 240
239 ctagacaaat tctacactga actctaccag cagctgaatg acctggaagc ctgtgtgata 300
240 cagggggtgg ggttgacaga gactcccctg atgaaggagg actccattct ggctgtgagg 360

```

## RAW SEQUENCE LISTING

DATE: 06/12/2003

PATENT APPLICATION: US/09/915,873A

TIME: 14:30:08

Input Set : A:\Seqlist.txt

Output Set: N:\CRF4\06122003\I915873A.raw

```

241 aaatacttcc aaagaatcac tctctatctg aaagagaaga aatacagccc ttgtgcctgg 420
242 gaggttgtca gacagaaat catgagatct ttttctttgt caacaaactt gcaagaaagt 480
243 ttaagaagta aggaatga                                     498
245 <210> SEQ ID NO: 11
246 <211> LENGTH: 569
247 <212> TYPE: DNA
248 <213> ORGANISM: Homo sapiens
250 <400> SEQUENCE: 11
251 atggccttga cctttgcttt actggtggcc ctctgtgtgc tcagctgcaa gtcaagctgc 60
252 tctgtgggct gtgatctgcc tcaaaccac agcctgggta gcaggaggac cttgatgctc 120
253 ctggcacaga tgaggagaat ctctcttttc tctgtcttga aggacagaca tgactttgga 180
254 tttccccagg aggagtttgg caaccagttc caaaaggctg aaaccatccc tgtcctccat 240
255 gagatgatcc agcagatctt caatctcttc agcacaaagg actcatctgc tgcttgggat 300
256 gagacctcc tagacaaatt ctacactgaa ctctaccagc agctgaatga cctggaagcc 360
257 tgtgtgatac agggggtggg ggtgacagag actcccctga tgaaggagga ctccattctg 420
258 gctgtgagga aatacttcca aagaatcact ctctatctga aagagaagaa atacagccct 480
259 tgtgcctggg aggttgtcag agcagaaatc atgagatctt tttctttgtc aacaaacttg 540
260 caagaaagtt taagaagtaa ggaatgaaa                                     569
262 <210> SEQ ID NO: 12
263 <211> LENGTH: 62
264 <212> TYPE: DNA
265 <213> ORGANISM: Arabidopsis thaliana
267 <400> SEQUENCE: 12
268 atgaagacta atctttttct ctttctcatc ttttcaactc tcctatcatt atcctcggcc 60
269 ga                                     62

```

VERIFICATION SUMMARY

DATE: 06/12/2003

PATENT APPLICATION: US/09/915,873A

TIME: 14:30:09

Input Set : A:\Seqlist.txt

Output Set: N:\CRF4\06122003\I915873A.raw

L:53 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!

L:56 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2